Contents

Chapter 1 • Points of View

Each thing we see hides something else we want to see.—René Magritte

Whether you are preparing a 10-page pamphlet or a 300-page book, you can view the process of creating and producing an electronic document in many different ways. The better you understand all these points of view, the more effective you will be in choosing and using the available software tools.

Each software tool presents a particular conceptual model of the publishing process. This *philosophical* point of view greatly influences the functionality and usability of the software.

Some systems are *page oriented.* Others focus on the entire document. Some are WYSIWYG (what you see is what you get). Still others are *batch oriented*. Learning and using the publishing tools are easier if you are aware of the philosophy—the point of view—that a system supports.

One way to grasp the value of new technologies is to create a metaphor. The iconic user interface of the Macintosh is known as a **desktop**. The use of printing software and hardware on this metaphorical desktop is known as **desktop publishing**. It brings to mind miniature Gutenberg presses right at your fingertips. All sorts of desktop metaphors have been created: desktop machining, desktop forgery, desktop prepress, and so on.

The newer technologies of electronic publishing also need new metaphors to cover the issues of document processing, electronic distribution, archival storage, and so on.

Where is that catchy phrase?

Aldus Magazine, a magazine about desktop publishing, graphic design, and other document-oriented issues, had a contest and asked its readers to "create one or more terms that express the essence of where we are going with computers and communications." In the end, the magazine did not choose any one phrase as a winner. Instead, it listed many of the interesting phrases suggested. The entire attempt illustrates the difficulties of creating new metaphors.[1]

AutoPublishing
Cinematic Publishing
Communication Design
Communication Network
CompCom
Computer Aided Publishing
Computer Aided Thinking
Computer Expression
Concept Communication
Creative Interlink
Creative Resourcing
Design Interchange
Design Link
Design Transmit
Device Independent Publishing
Digital Communications Management
Digital Design
Digital Extensions
Digital Media
Electronic Expressionism
Electronic Multidimensional Interlink
Expressions Enhancement
Fingertip Communication
Human Creations Technology
Human Techniscreening
Idea Translation
Instant Design Communication
Instant Link
Intermedia Transmutation
Media Toolkit
Multipublish
Multitechnical
Neocommunication
Neoscribe
Network Design
Network Talk
Perspective Composition
Publications Processing
Sensory Perception
Universal Media Link
Varimedia Translation Nexus

The multifaceted world of electronic publishing needs a catchy phrase to describe it. The many points of view that you can use to examine electronic publishing are necessary because there is no single satisfying metaphor.

The term **electronic publishing** means different things to different people. Many of the standards discussed in this book open up other possibilities such as hypertext, on-line information browsers, and so on; these applications, as well as databases, CD-ROMs, and other electronic repositories, are beyond the scope of this book. The majority of this book is concerned with electronic documents that are to be printed.

In this chapter we will examine many approaches to looking at electronic documents—points of view. The views we will examine are (1) Visual and Logical views, (2) The Design Point of View, (3) Communications Views, (4) The Engineered View, (5) The Database View, and (6) Specialized Views. We will look at how the creation of electronic documents is influenced by each point of view.

1 • 1 Visual and Logical Views

Documents have many components—characters, words, paragraphs, chapter headings, sections, and subsections. We can examine each component in two complementary ways, the visual and the logical.

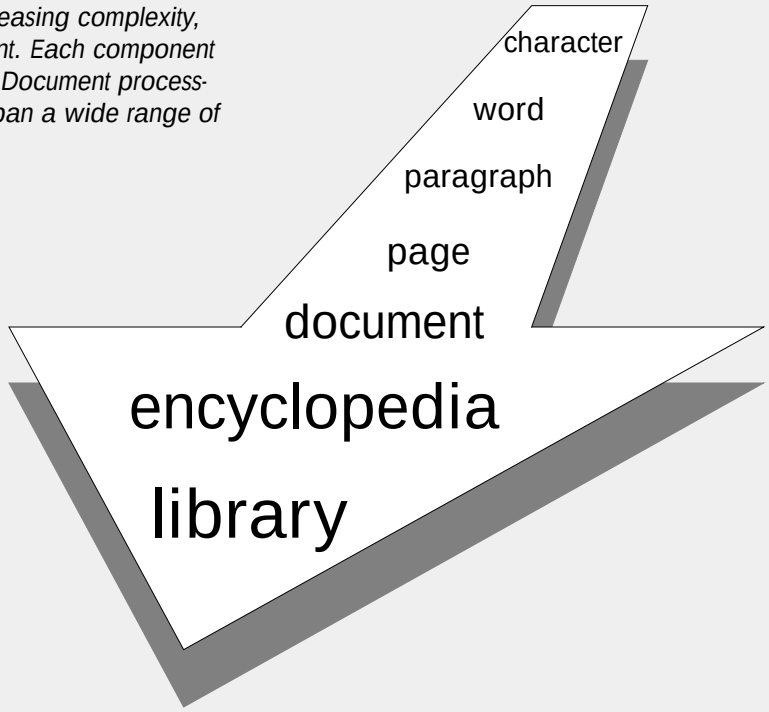
The logical aspect of a component refers to its semantically meaningful part, such as the fact that a collection of characters is a word that can be checked for spelling, or that a chapter is divided into sections. The visual aspect of a document component refers to the size, position, and fonts used to form its physical appearance.The visual components of document elements will be discussed further in *Section 1 • 2 The Design Point of View*.

In this section, we will examine document components of increasing complexity, starting with the character and progressing through to an entire

library. Each document component has a visual aspect and logical aspect. Some lean more toward one than the other.

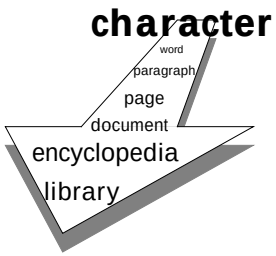
Putting these document components on a scale, starting with the simplest and moving to the most complex provides us with a useful frame of reference in which to discuss these issues.

A Document Component Scale. These are the components, shown in order of increasing complexity, that form an electronic document. Each component has logical and visual aspects. Document processing systems and technologies span a wide range of document components.



You can manipulate each item on this scale using software tools. Of course, some tools cover several items on the scale. The orientation of a particular tool—the point of view it supports—will probably be centered around one particular item. In the following sections we will go through the scale by examining each document component.

1 • 1 • 1 Character Manipulation



The first level of our document scale is the character. Characters, as logical meaningful entities, have values that are represented in the computer according to well–known and established character codes. ASCII is the best known and established character encoding. Character codes are the fundamental representation of text.

Normally you don't have to be concerned about the character code used in your particular system. However, when you want to interchange to other systems, the character code may become a problem. In particular, interchange with systems in countries that use other character codes must take these codes into account. Many Asian languages require other character codes, which are necessary to support hundreds or even thousands of characters (for example, Japanese). *Localization* is the process of taking software written for one system and porting it to another system that uses another language and possibly another character code.

Also on the logical (as opposed to visual) side of the discussion is the ability to associate attributes or **tags** with individual characters. Essentially, tags are names you can associate with characters for whatever purpose you like. For example, the FrameMaker publishing system allows the definition of character tags. Each tag defines a particular font family, size, weight, and other properties, which can be applied to any character. These *tagged* characters may subsequently be manipulated as a group if necessary.

Named attributes or tags such as these provide a convenient mechanism for manipulating the visual appearance of characters throughout a document. You can also use them for other semantic purposes. For example, you could associate the name "placeHolder" with particular characters you wish to use temporarily. You can search for the tag "placeHolder" to locate the particular text. You can even print out a report listing all occurrences of the "placeHolder" tag and where they occur in the document, creating an automated list of remaining work.[2]

Is ASCII Dead?

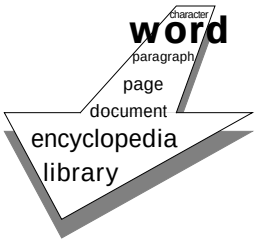
The need to print clear multilingual text in today's global computing community has made ASCII obsolete. The characters defined by ASCII are 7 bits wide, allowing only 128 characters. Many languages have extra characters with umlauts, accents, and so on. This has given rise to the new ASCII called ISO Latin 1 (ISO standard 8859:1), a clear text encoding for characters with 8 bits per character.

There's also Unicode, a 16-bit character representation created by a consortium of computer vendors. It is an attempt at representing ALL characters, including those from China, Japan, and Korea, where the languages have several thousand characters. This is a truly global attempt at character representation.[3]

For the visual side of characters many font manipulation tools are available. Tools to manipulate individual characters could be considered part of font definition software. If you want to change the appearance of all occurrences of the character T, for example, you use a font definition tool.

There are many more issues concerning the visual aspects of characters and fonts. Please see *Section 1 • 2 • 1 Fonts and Typography* later in this chapter for a discussion of these issues.

1 • 1 • 2 Words

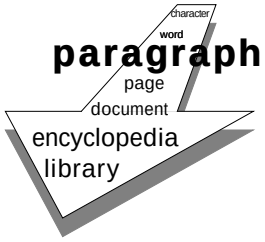


The act of writing takes place at the word level of our document element scale. Most of the discussion about writing is in *Section 1 • 3 Communications Views*, later in this chapter. Spelling checkers and grammatical aid systems are some of the electronic publishing tools that aid writing. The growing popularity of computer-assisted writing aids attests to their growing sophistication.

Another manipulation of words is automatic hyphenation. This is a manipulation of the logical or semantically meaningful aspects of words. Often, publishing systems allow the user to modify some variables to control the precise way automatic hyphenation is performed. For example, these could be variables to control the minimum and maximum number of characters before and after the hyphen. In addition, electronic publishing systems that support several languages must also have hyphenation dictionaries appropriate for each language. Hyphenation algorithms differ among systems.[4] The same document in two systems may not appear exactly the same, even if the fonts and page margins are identical, because the hyphens will break the words at different places. Hyphenation is part of the process of formatting and can hinder efforts to interchange documents with perfect fidelity.

1 • 1 • 3 Paragraphs with Tags and Styles

Moving up the complexity scale, we now come to the paragraph. One of the most powerful document processing tools is the ability to attach attributes, tags, or



styles to paragraphs. I use the term **tags** to refer to the logical aspect of paragraphs and **styles** to refer to the visual aspect of paragraphs.

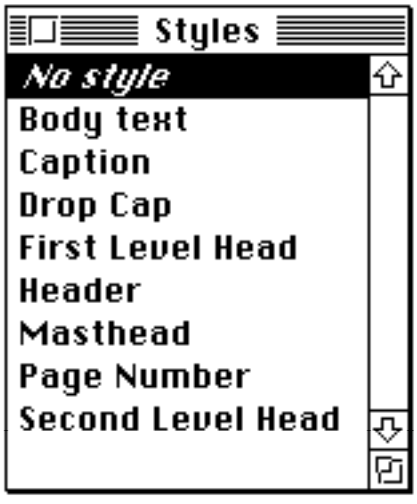
When writing, we generally treat the content and appearance of paragraphs uniformly. Individual paragraphs have the same margins and typefaces (they should *also* contain a coherent idea). Many software products treat the paragraph as an entity that can be manipulated as a unit.

When manipulating a paragraph, it is important to distinguish the logical aspects from the visual. The logical use of a paragraph tag might be to identify all chapter headings. The publishing system may support the intent of a document structure and not allow the creation of a chapter heading in the middle of a table. Identification of the logical structure of a document is one of the major features of formal document standards and is discussed in depth in the Document Standards chapter. (Please see *Section 3 • 3 SGML* in *Chapter 3 • Document Standards* for a discussion of document structure.)

Another logical use of tag names is the actual name itself. The name “Body text” conveys the meaning that the body copy in a document will be associated with the tag “Body text.” It is important to select meaningful tag names. Cryptic, “cutesy” names obscure the intent of the tag or style. Good names are vital. Spend the painful time creating good names that will be meaningful to others in your organization.

The development and use of a consistent set of paragraph tags can be of tremendous value. This task should be done at the start of any significant project. Visual consistency can be achieved by using the same tags in the same places. Just as important, changes can be applied to specific tags or styles in one place and then applied to the entire document. The concept of a style sheet is intended specifically to allow changes in one place to migrate to the rest of the document. Changes made to a style sheet can also be applied to many other documents, helping to automate and keep consistent all documents of a

Style Names from PageMaker

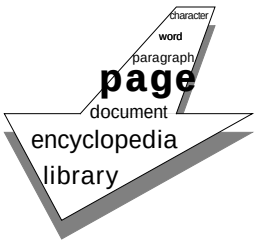


project or organization. Coherent tag names allow the logical aspects of the document to guide the visual appearance.

Let's turn now to the next level in the document element scale, the page.

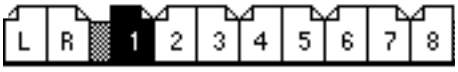
1 • 1 • 4 The Page

Contrary to the other document components, the page is purely visual and has no meaningful logical aspects. Pages are the physical spaces in which textual content appears. Page sizes can be altered and documents can be reprinted in different sizes and formats for on-line browsing and so on, with no effect on the content. Pages do not have any logical aspects other than their very existence. They represent a canvas upon which the content is painted.



From a visual point of view, pages provide a place for a number of items. Headers, footers, body text, and page numbers are some of these items. They are placed on pages, in a consistent position throughout the document. The positioning of these items is primarily a matter of design. In addition to design, however, there are computational factors. Some of the page-specific items, such as the page numbers, running headers, and running footers, can be computed or extracted from the text. The content of these items can be changed, based on the specifics of the page.

Although pages have a specific size that is rarely changed, paying attention to the size is sometime crucial. Many systems support specific page sizes implicitly. This implicit assumption can cause a nasty problem if you need to interchange documents with an organization that uses a different standard page size than your own organization. This would probably happen when a U.S. organization exchanges documents with an organization based in Europe. U.S. standard page sizes (8.5 x 11 inches) are different than the ISO A4 (8.25 x 11.75 inches) size used in Europe. The document will probably not print correctly unless you adjust for page size.



An iconic view of left and right handedness from PageMaker.

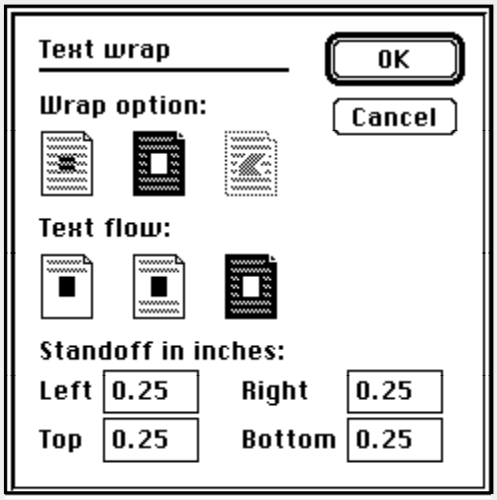
The layout and overall design of components such as text, graphics, and illustration are best manipulated in a **page layout program**. The quintessential example of this type of software is Aldus PageMaker. One of the keys to PageMaker's success is that this software speaks the language of designers. Its presents the user with a simulation of a pasteboard (an underlying grid for creating the proportions and overall structure of the document), a commonly used graphic design tool.

One distinction that must be applied only to pages is *handedness*—whether the content of a page is to appear on the right-, or left-hand side of the printed document. Margins, columns, headers, footers, and page number positions are sometimes shifted on the page, depending on whether they are to appear on a left or right-hand page. The more powerful electronic publishing systems provide tools to control handedness: for example, the ability to force the start of each document (for example, chapters) on a right hand page.

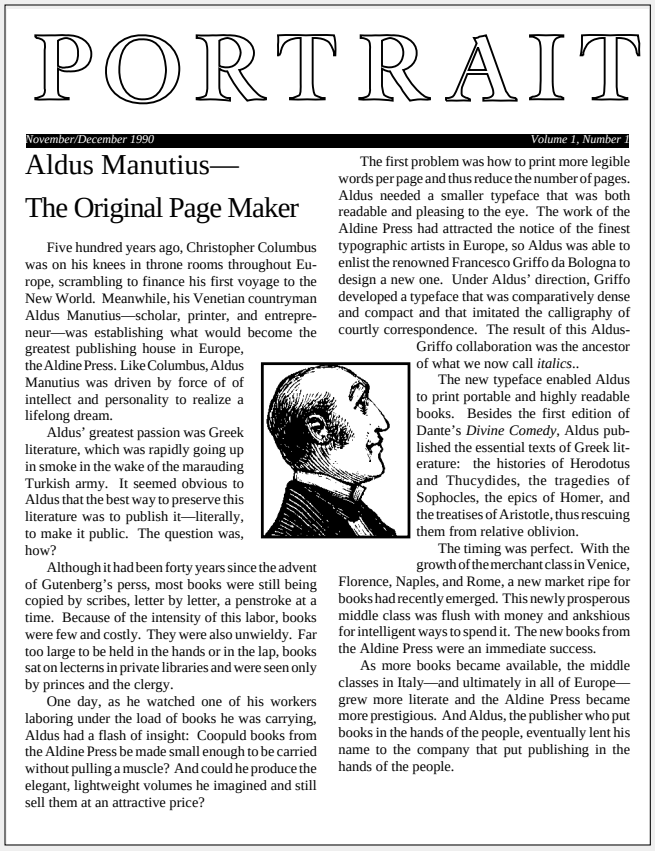
Text flow is yet another term that really crosses the boundary from a page to a document. Newspaper articles leave pointers to the connecting text, such as "*see Bozos column 5, page 22*". These pointers tell the reader where the text is continued. The visual shape of these flows is either rectangular or follows the shape of graphic elements. Page layout or page makeup programs such as QuarkXPress and Aldus Pagemaker provide tools that allow text flows to travel automatically around graphic elements.

Flowing and Wrapping Text

Text flowing is a technique used in many page makeup and publishing systems. Virtually all page layout programs allow the text to flow from one column to another. Some systems have facilities to easily wrap text around graphic illustrations.[5]



PageMaker 4.0 dialog box for the control of text wrapping.



PORTRAIT

November/December 1990 *Volume 1, Number 1*

Aldus Manutius—

The Original Page Maker

Five hundred years ago, Christopher Columbus was on his knees in throne rooms throughout Europe, scrambling to finance his first voyage to the New World. Meanwhile, his Venetian countryman Aldus Manutius—scholar, printer, and entrepreneur—was establishing what would become the greatest publishing house in Europe, the Aldine Press. Like Columbus, Aldus Manutius was driven by force of of intellect and personality to realize a lifelong dream.

Aldus' greatest passion was Greek literature, which was rapidly going up in smoke in the wake of the marauding Turkish army. It seemed obvious to Aldus that the best way to preserve this literature was to publish it—literally, to make it public. The question was, how?

Although it had been forty years since the advent of Gutenberg's perss, most books were still being copied by scribes, letter by letter, a penstroke at a time. Because of the intensity of this labor, books were few and costly. They were also unwieldy. Far too large to be held in the hands or in the lap, books sat on lecterns in private libraries and were seen only by princes and the clergy.

One day, as he watched one of his workers laboring under the load of books he was carrying, Aldus had a flash of insight: Coopuld books from the Aldine Press be made small enough to be carried without pulling a muscle? And could he produce the elegant, lightweight volumes he imagined and still sell them at an attractive price?

The first problem was how to print more legible words per page and thus reduce the number of pages. Aldus needed a smaller typeface that was both readable and pleasing to the eye. The work of the Aldine Press had attracted the notice of the finest typographic artists in Europe, so Aldus was able to enlist the renowned Francesco Griffo da Bologna to design a new one. Under Aldus' direction, Griffo developed a typeface that was comparatively dense and compact and that imitated the calligraphy of courtly correspondence. The result of this Aldus-Griffo collaboration was the ancestor of what we now call *italics*..

The new typeface enabled Aldus to print portable and highly readable books. Besides the first edition of Dante's *Divine Comedy*, Aldus published the essential texts of Greek literature: the histories of Herodotus and Thucydides, the tragedies of Sophocles, the epics of Homer, and the treatises of Aristotle, thus rescuing them from relative oblivion.

The timing was perfect. With the growth of the merchant class in Venice, Florence, Naples, and Rome, a new market ripe for books had recently emerged. This newly prosperous middle class was flush with money and ankshious for intelligent ways to spend it. The new books from the Aldine Press were an immediate success.

As more books became available, the middle classes in Italy—and ultimately in all of Europe—grew more literate and the Aldine Press became more prestigious. And Aldus, the publisher who put books in the hands of the people, eventually lent his name to the company that put publishing in the hands of the people.

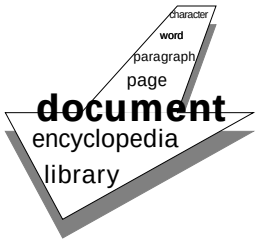
Page with text flow wrapping around graphic.

Frames are another frequently encountered term with a strong relationship to the page. In a sense, a frame is a subdivision of a page. It is an invisible boundary in which content appears, just like a page. Frames, however, are not physical things; they are areas that can be manipulated while using the pub-

lishing system. Text can flow automatically from one particular frame to another. Ventura Publisher[6] and FrameMaker[7] both use this concept.

Last but not least, Interleaf 5[8] generalizes many of the aspects of a page in a feature known as a microdocument. Microdocuments are "little" documents, inserts embedded in the pages of other documents, that can independently retain stylistic characteristics. All the styles associated with a particular document can be retained intact with microdocuments. However, the microdocument can be no larger than a page; hence, its name.

1 • 1 • 5 The Document



The document in its entirety is the next stop in our analysis of document components. From a visual point of view, the document is a physical object with a particular design. From a logical point of view, the document is composed of a certain structure. The visual design and construction of documents[9] is a topic beyond the scope of this book. However, electronic publishing systems can play an essential role in the manipulation of the logical aspects of a document.

The logical structure of a document is an important characteristic of the document. We can use that structure as a framework to evaluate document processing tools. Some questions to ask in determining the suitability of a particular publishing system are:

Can the system automatically generate a table of contents?

Can the system generate lists of various elements such as tables and figures?

What kind of graphics can be integrated easily with the text?

How robust are the indexing capabilities, if any?

Is there good bibliographic and cross-referencing support?

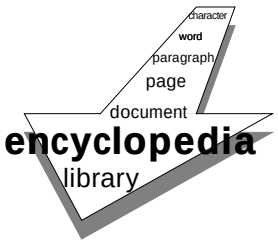
Technical publications, in particular, need robust document-oriented tools. The more automated tools, the better. It is essential that the publishing system

provide support for automatic section numbering, running headers and footers, styles or tags, and change control. In addition, support for global changes—changes to many files that are part of a large document—is a major time saver.

Several publishing systems present the user with the idea of a book.[10] The concept of a book is used as an organizational tool. Books are made up of collections of files. If a change is made to the book, then the change is actually made to all the files that make up the book. If your publishing projects routinely deal with hundreds of files, this type of support will be an important requirement for any publishing system.

As the sheer size of the document grows, we start to see a significant distinction between WYSIWYG (what you see is what you get) and batch–oriented systems. Often you don't want to see extensive, repetitive, massive changes. If you are forced into too many hand manipulations, the publishing system may be unwieldy for the particular publishing application. The higher-end publishing systems try to balance WYSIWYG capabilities with the often awkward and complicated commands of a batch–oriented system. Please see *Section 2 • 1 Types of Document Processors* in *Chapter 2 • Form and Function of Document Processors* for a more through discussion of WYSIWYG versus batch document processing.

1 • 1 • 6 The Encyclopedia: A Multivolume Document



When we discuss the multivolume or encyclopedic scale of documents, our focus shifts from document manipulation to the concept of a data repository. Manipulation of large quantities of related material is one of the strengths of batch–oriented document processing systems. Off–line automated processing is a virtual requirement for this scale of manipulation.

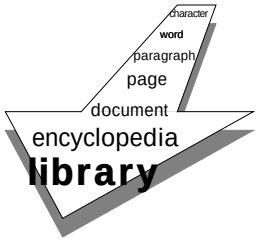
This level, in our document component scale, also represents the highest point at which a collection of documents is part of a coherent whole. Representative examples of documents in this level are the many manuals of an operating system, the volumes of an encyclopedia, and the maintenance manuals for a jet engine.

Interleaf 5 is a good example of a system with capabilities at this level. It uses the concept of a cabinet, which contains collections of other documents.

Only when a publishing system supports the manipulation of multiple volumes as a unit is the multivolume category qualitatively different from the previous category. The large volume of data and high capacities required for such manipulations are supported only by the high–end publishing systems.

Again, Interleaf 5 supports different types of style sheets, one that can be applied to individual documents and a master style sheet that is used to modify other style sheets called the master style sheet. Master style sheets are an important feature when massive and consistent changes are required. The batch–oriented document processing systems such as `troff` and `TeX` (see *Section 2 • 1 • 2 Batch Characteristics* in *Chapter 2 • Form and Function of Document Processors*) are also effective at working with massive amounts of material. Automated scripts can be created and documents processed without human intervention. In general, however, skilled technical users must create these scripts; they require a different type of staff than the turnkey (but more expensive) systems.

1 • 1 • 7 Library



A library, the final level in our document component scale, is discussed here because it relates to the topic of text retrieval. When maintaining or creating a library of documents or other large archival collections of documents, the technical issues are primarily ones of access. Finding information quickly and easily is the primary issue.

The most important area in which to address these issues is that of classification. Classification and searching systems are integral parts of library science. A good classification system enables users to locate the information they desire and aids in the management of the documents. After all, if you can't find the information you need, when you need it, you may as well not have it at all. One area where document processing and searching systems intersect is of full–text searching.

Full–text searching is the ability to search for any word in an entire collection of documents. The searching is usually accomplished through the use of a **document browser**. The emphasis in full–text searching is on speed at the sacrifice of space. It is not unusual for the indexes used to locate the text to take up as much space as the text itself. The combination of a good document browser and full–text searching really makes the entire field of electronic books a useful practical commodity, rather than an interesting toy.

Full–text retrieval engines are widely used in the creation of systems that manage large quantities of text. Retrieval engines are becoming quite prevalent in the CD-ROM industry[11] and are a key technology to enable access to a library full of information. The large capacity of CD-ROMs is an ideal complement to the large space requirements of full–text retrieval systems.

Text retrieval is a complex field, which is growing in importance as the world gets interconnected ever more tightly with networks.[12] The increased capacity of low–cost storage devices like CD-ROMs is also a major factor in text retrieval, because entire databases can be put online right at your very own PC. For more information on text retrieval, please see *Section 6 • 3 • 2 Text Retrieval* in *Chapter 6 • Document Management*.

That about wraps up our analysis of document components. One additional step to think about is **global networked information**. The rapidly solidifying collection of information, accessible via networks, may quite realistically form a global library. The technical barriers to such a fantasy are quickly disappearing. Cellular modems with portable laptop workstations are a reality. Only the legal concerns (which are not minor) of intellectual property rights, copyrights, and patent law remain as murky unknowns. For a more through discussion of the possibilities of networks see *Section 5 • 4 Electronic Distribution* in *Chapter 5 • Using Standards*.

Everything Old Is New Again

The Dewey Decimal System is a classification system, widely used in libraries. It has the interesting property of infinite expansion. Similarly, Ted Nelson, the man who coined the term hypertext and one of the pioneers of hypertext systems, has taken the idea of variable precision addressing for his Xanadu system. He calls this address scheme *tumblers*, in which any place can be expanded. It's sort of like an outline processor for computer addresses. "The tumbler space is an accordion-like master address space..."[13]

1 • 2 The Design Point of View

Design is another point of view that must be considered as we examine ways of approaching the document-creation process. The way a document is visually presented, how it grabs the audience visually, is a critical factor in the overall perception of a document. After all, the end product is an object to be viewed. The aesthetic components that make up the pages, fonts, layout, and color all contribute to the overall goal of producing a document that communicates ideas clearly. A through treatment of document design is beyond the scope of this book, but for a list of good books see *Section Publications* in the *Appendix Resources*. The rest of this section will introduce some of the basics of document design and other topics with strong relationships to document processing.[14]

1 • 2 • 1 Fonts and Typography

Typography is to writing what a soundtrack is to a motion picture.
—Jonathan Hoefler

Open up any computer magazine about desktop publishing and you will see many ads for fonts and font-manipulation software. It may seem that the world has gone a little font crazy. Fonts specifically and typography in general are extremely important, and this variety of type is natural.

In some sense, typography is something that is so obvious, so visible, and all-encompassing that most people simply don't notice it. However, it is precisely because typography is so pervasive that it is so important.[15] Fonts are not simply the shape of letters for creating words; they are letterforms with carefully designed shapes with subtle differences that relate to each other and that combine to make a pure visual statement.

Appropriate Fonts

Some fonts are simply more appropriate than others for particular uses. The figures below illustrate inappropriate and appropriate font usage. You decide.

The Declaration of Independence

When in the course of human events, it becomes necessary for one people to dissolve the political bands which have connected them with another, and to assume among the powers of the earth, the separate and equal station to which the laws of Nature and of Nature's God entitle them, a decent respect to the opinions of mankind requires that they should declare the causes which impel them to the separation.

We hold these truths to be self-evident, that all men are created equal, that they are endowed by their Creator with certain unalienable rights, that among these are life, liberty and the pursuit of happiness. That to secure these rights, governments are instituted among men, deriving their just powers from consent of the governed,—That whenever any form of government becomes de-

The Declaration of Independence

When in the course of human events, it becomes necessary for one people to dissolve the political bands which have connected them with another, and to assume among the powers of the earth, the separate and equal station to which the laws of Nature and of Nature's God entitle them, a decent respect to the opinions of mankind requires that they should declare the causes which impel them to the separation.

We hold these truths to be self-evident, that all men are created equal, that they are endowed by their Creator with certain unalienable rights, that among these are life, liberty and the pursuit of happiness. That to secure these rights, governments are instituted among men, deriving their just powers from consent of the governed,—That whenever any form of government becomes destructive of these ends, it is the right of the people to alter or to abolish it, and to institute new government, laying its foundation on such principles and organizing its powers in

Some software tools pay more attention than others to the role of fonts and typography. Depending on your specific needs, these tools may or may not be important. However an awareness of the crucial factors can only help when judging the capabilities of a particular tool. In general, page makeup and page layout programs have much more flexible typographic features than their batch–oriented counterparts. The WYSIWYG nature of page makeup systems is more suitable to adhoc design and experimentation.

Serif vs. Sans Serif Fonts

There are two broad categories of fonts: serif and sans serif. Serif fonts are fonts with finishing strokes at the ends of the main strokes of the letters. Many serif fonts exhibit calligraphic qualities in their main strokes. Sans serif fonts do not have the finishing strokes (serifs) and are often considered more modern. Overall, serif fonts are easier to read. As a general (albeit very oversimplified) rule, sans serif fonts are better for headings but not for main body text.

Serif examples

Originally letters were adaptations of natural forms employed in picture writing, but by a process of evolution [actually degradation] they have become arbitrary signs with little *resemblance to the symbols from which they are derived. These arbitrary shapes* have passed through their periods of uncertainty and change; they have a long history and manifold associations; they are classics, and should not be tampered with, except within limits which a just discretions may allow. — Fredrick Goudy
The Alphabet and Elements of Lettering

Sans serif examples

The test which a well-formed letter must meet is, that nothing in it shall present the appearance of being an afterthought—that every detail shall at least seem to have been foreseen from the start; and, when letters are used in combinations to form words and sentences, that no one of them shall stand out from its fellows or draw attention to itself at the expense of those with which it is associated. — Fredrick Goudy
The Alphabet and Elements of Lettering

If you are faced with selecting a font, it is important to consider the number of variations available in a font family. Some font families have more than a

dozen variations. This flexibility can only make the designer's job easier. Using several variations within a single font family is almost always aesthetically safer than mixing arbitrary fonts.

Variations on a Font

Some font families offer many variations from which to choose. Sometimes specialized variations exist for specific decorative uses such as titling, fractions, ligatures, and "old style" looks.

Futura Light
Futura Book
Futura Bold
Futura Heavy
Futura Extra Bold
Futura Light Oblique
Futura Book Oblique
Futura Bold Oblique
Futura Heavy Oblique
Futura Extra Bold Oblique
Futura Condensed Light
Futura Condensed Book
Futura Condensed Bold
Futura Condensed Extra Bold
Futura Condensed Light Oblique
Futura Condensed Book Oblique
Futura Condensed Bold Oblique
Futura Condensed Extra Bold Oblique

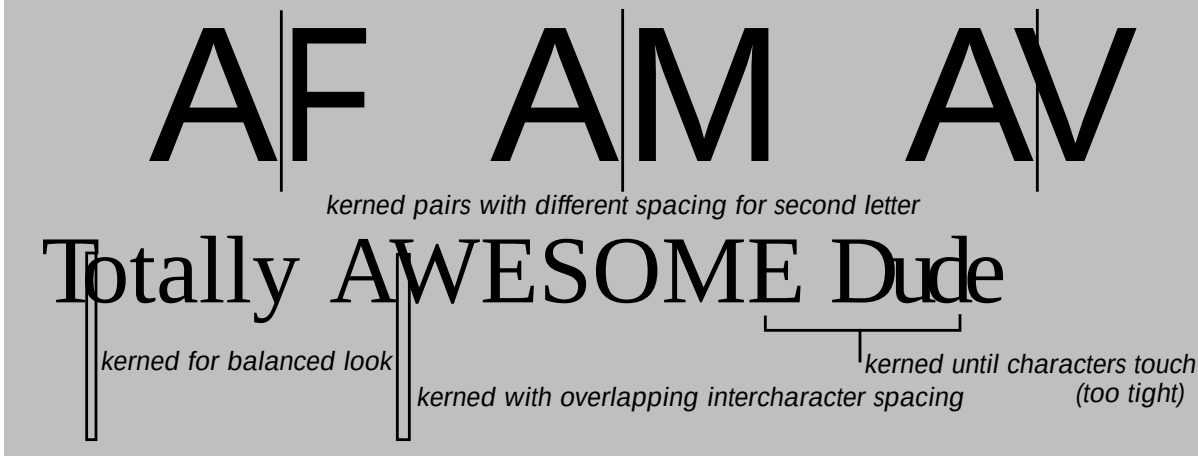
Adobe Garamond
Adobe Garamond Semibold
Adobe Garamond Bold
Adobe Garamond Italic
Adobe Garamond Semibold Italic
Adobe Garamond Bold Italic
ADOBE GARAMOND TITLING CAPITALS
ADOBE GARAMOND EXPERT
ADOBE GARAMOND EXPERT SEMIBOLD
0123456789¼½⅛⅜⅝¾⅔⅓¢ffiffl — *Adobe Garamond Expert Bold*
0123456789¼½⅛⅜⅝¾⅔⅓¢ffiffl — *Adobe Garamond Expert Italic*
0123456789¼½⅛⅜⅝¾⅔⅓¢ffiffl
└— *Adobe Garamond Expert Semibold Italic*
ADOBE GARAMOND ALTERNATE ITALIC
a & tenrttzQ&S — *Adobe Garamond Alternate*

Many tools are available for font manipulation. These tools allow precise adjustments of kerning tables (the spacing between letters), the creation of new letterforms, the extraction of outlines, distortions, and so on. One important reason that such a variety of detailed tools exist is that font design has such an important impact on the document as a whole. Letterforms are a key ingredient in a document, and designers use them as the raw material to be  by these tools.

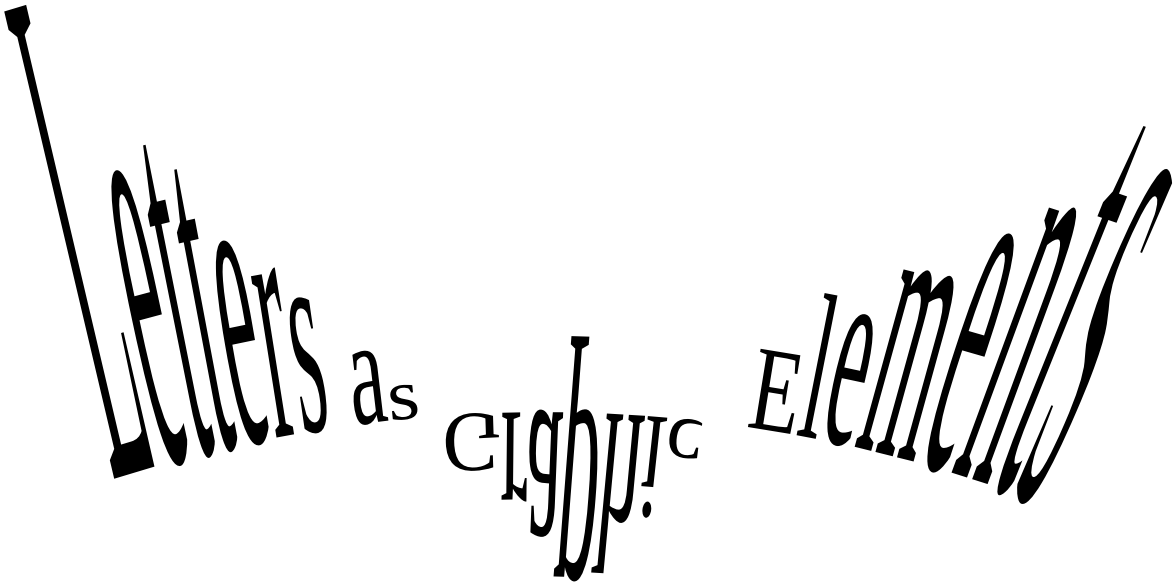
Of course, it's important not to get carried away with these tools.

Kerning

To kern letters is to adjust the exact amount of spacing between characters. The goal is to visually balance a particular piece of text. A kerning table is one of the pieces of data that defines a complete font. A full kerning table defines the spacing between pairs of letters, not simply the amount of space after each individual letter.



Individual characters may also be used as graphic components. The line between font manipulation and graphic illustration can blur quite easily.



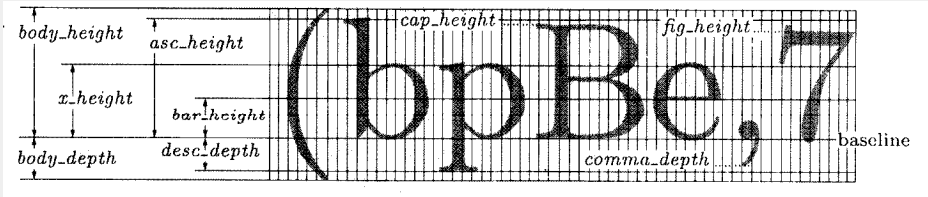
The many software tools available for font manipulation allow such a wide variety of choices that the traditional letterform is no longer sacred. Characters used as illustrative elements bring us back to the age of illustrated manuscripts filled with carefully crafted characters. There is of course the added danger of "font junk," the use and abuse of font manipulation tools, by the amateur.

Another somewhat obtuse but powerful character manipulation system is the METAFONT language.[16] METAFONT is a precise mathematical description of fonts; in many ways it models the way ink is placed on paper by a pen. METAFONT is the creation of Donald Knuth—the same man that brought you TeX (see *Section 2 • 1 Types of Document Processors*). METAFONT is a language for describing characters in excruciatingly precise terms. After creating or modifying a description, the system chews away on the "code" and spits out a new font. These fonts can than be used by TeX, turning this interesting academic exercise into a practical and useful tool.

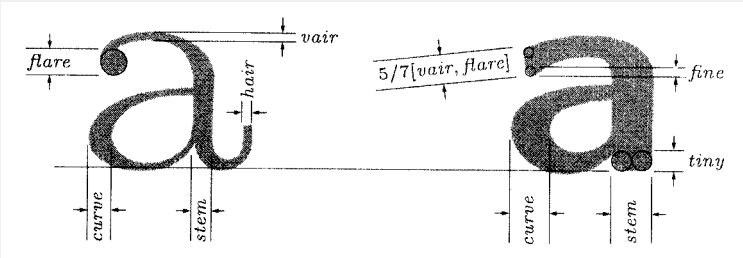
Some METAFONT Parameters

These are some of the over 60 parameters manipulable in the Computer Modern typeface, created with METAFONT.

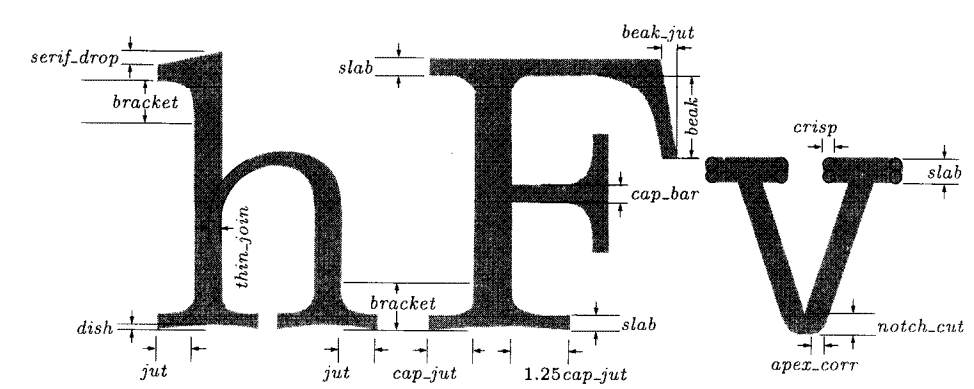
vertical parameters for height and depth manipulation



darkness or weight manipulation parameters



parameters for controlling serifs and arms



1 • 2 • 2 Layout and Composition

The placement of the various components of a document on a page is the **layout**. Document layout and composition are critical pieces of the design puzzle. Unfortunately, the only help electronic publishing tools have to offer is assistance in the use of templates. Tools that aid in the overall layout and structural composition of documents exist only in research laboratories. Automated aids for global design features such as overall balance, proper use of white space, and so on, do not exist as product features.

Typical document processing systems have **style sheets** or **master pages**, which define a particular visual layout. The visual layout of document elements on the style sheets can be applied to the entire document. The number of master pages and the flexibility in working with them are important capabilities of a document processing system. Often, global changes to a document are accomplished using these types of pages or styles. Careful use of master pages and style sheets is a significant help in the management of overall document consistency. (For a more through discussion of document management issues, please see *Chapter 6 • Document Management*.)

In the future it may be possible to have design "helpers" in much the same way that grammar checkers now exist. Such suggestions are not pure fantasy. We are already starting to see the application of image-recognition systems in the pen-based portable computer field. Users can create rough sketches and the system cleans up the drawing.[17] Image recognition is being taken a step farther with the concepts of **shape grammars**.[18] In the architecture and computer graphics domains, shape grammars have been used to create simulated buildings in the style of Frank Lloyd Wright[19] and paintings by Kandinsky.[20] The concept is to create a grammar, a language, from a set of shapes and the allowable operations upon those shapes. Many interesting grammars have been created to describe the styles of architects and artists.

1 • 3 Communications Views

When the writer becomes the center of his attention, he becomes a nudnik. And a nudnik who believes he's profound is even worse than just a plain nudnik.
—Isaac Bashevis Singer

First and foremost, a document is a tool to communicate information. The type of information will affect the type of communications. Some different information types are entertainment, reference, scanning, mandatory versus optional, sales, friendly, and formal. Each information type has customary visual conventions. Used poorly or too often, they will cause your document to look like just another piece of garbage. Used judiciously and with imagination, they can be a valuable aid.

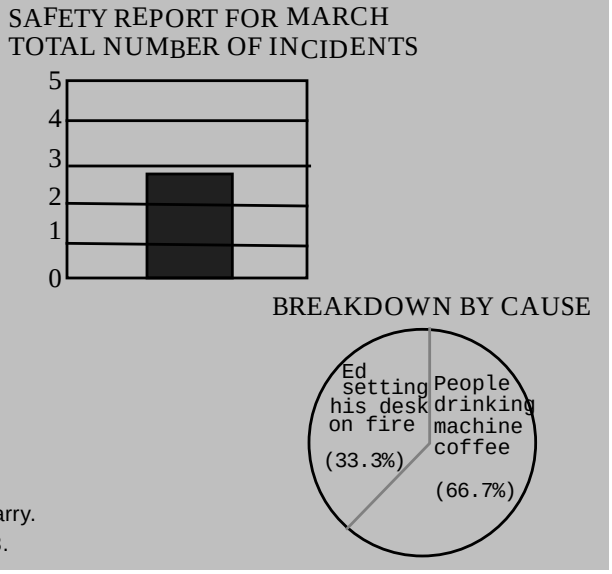
Ultimately, the content expressed in the document is what really matters. If the reader understands the content, your communication was successful.

HOW TO USE COMPUTER-GENERATED PIE CHARTS AND BAR GRAPHS TO MAKE ABSTRACT CONCEPTS UNDERSTANDABLE TO MORONS LIKE YOUR BOSS

Let's say you have to write a Safety Report. The old-fashioned, pre-computer way to do this would be something like this:

```
In March, we had two people who got
sick because they forgot and drank
coffee from the vending machine.
Also, Ed Sparge set fire to his desk
again. Ed has promised that from
now on he will put his cigar out
before he dozes off.
```

But now using the graphics capability on your computer, you can produce a visually arresting and easy-to-understand report like this:



Often the main trick is getting the reader to pay attention. Look at some of your junk mail; innumerable attention–getting devices will come into view. Colored stamps, fake telegrams, pop-ups, personalized names, metallized envelopes, and more are all attention grabbers.

In the domain of electronic documents, clip art collections of all sorts can help you draw attention to your documents. Clip art collections with all sorts of specialty images (see *Section Clip Art* in the *Appendix Resources*) from military symbols to biological parts to cartoons, can help get the message to the reader. Clip art and unimaginative attention–getting devices can cut both ways, however.

Looks Aren't Everything

If you owned stock in these companies which Annual Report would you rather receive?

The Big Bad Beautiful Corporation A N N U A L R E P O R T

Earnings throughout 1992 were slow due to the progressive erosion of market share. The downsizing effort will continue until quarterly cash flow improves.

	Year ending October 31	
	1993	**1994**
	(in thousands)	
Net dividends to shareholders from:		
Investment income	$ 40,111	$ 5,432
Realized gain on investment transactions	$ 12,555	$ -1,045
Total dividends	$ 43,111	$ 0
Net Income (LOSS):		
Beginning of year	$644,972	$(200,369)
End of year	$120,804	$(576,972)

The Bigger Baddest Inc. Annual Report

This was it! The year we finally broke even and started to make a profit! The R&D finally paid off. Market share is growing. Happy dividends.

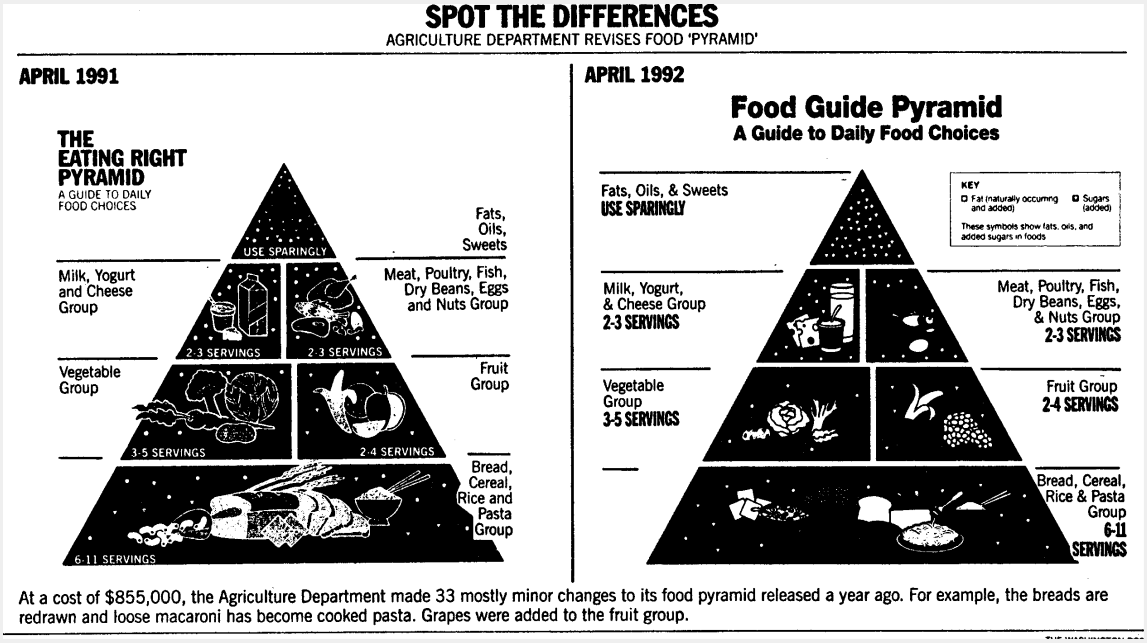
	Year ending, October 31	
	1993	1994
Net dividends to shareholders from:		
Investment income	2,000	12,342
Realized gain on investment transactions	1,322	6,578
Total dividends	1,778	4,223
Net income:		
Beginning of year	-31,000	1,332
End of year	-2,221	31,233

Customizing the content of an article for a particular audience is a good way of improving communication. Of course, doing this is extremely difficult for large–volume publications, such as newspapers and magazines. One interesting technique used by the *Washington Post* (and others) is called **zoning**. The *Post* has a column called Dr. Gridlock that describes the trials and tribulations of travel in the Washington, D.C., area. The content of this column is modified for specific areas by the use of readers' addresses via delivery zones.

SCIENCE, POLITICS, and FOOD PYRAMID GRAPHICS

Although design doesn't mean everything, it can have important and even political impact. For instance, take the case of the food pyramid.

In April 1991 the U.S. Department of Agriculture (USDA) was going to publish a replacement of the basic four food groups wheel, a staple of classrooms since the 1950s. The idea was to increase the importance of grains, fruits, and vegetables and reduce the importance of meat and dairy products, following good nutritional practices. As you might imagine, the beef and dairy lobbyists were not too happy about this turn of events. After a great deal of criticism, publication of the pyramid was halted. According to one nutritionist angered by the USDA reversal, "It was the visual that made the impact. That's what upset people; it clearly showed you should not have as much meats and dairy products as you should grains, fruits, and vegetables—which is the truth."[21]



One year later (and $855,000 more), the USDA unveiled a refined pyramid and had more data supporting its case. In the end, good science won out and the lobbyists had to live with the design of the food pyramid.[22] A final note-: in October 1992 as I was completing this book my kids brought home a lunch menu from their school. On the bottom of the menu was an explanation of good nutrition accompanied by—the food pyramid. I suppose the pyramid has become a new classroom staple.

Reprinted with special permission of King Features Syndicate

1 • 3 • 1 Aid for Grammarless Writers

A man's grammar, like Caesar's wife, must not only be pure, but above suspicion of impurity.— Edgar Allan Poe

As we examine ways in which technology can help in the communication of ideas, publishing systems can provide a number of tools to aid grammar. At times the technology of word processing and desktop publishing systems is more fun than writing. Integrated graphics with text, WYSIWYG displays, and font manipulations can divert the writer from the task at hand, communications. In an article titled *"Does Technology Contribute to Bad Writing? Perhaps It Might Probably Could—Or NOT,"* Michael Schrage, a columnist for the *Los Angeles Times*, printed in *The Washington Post*, comments:

Indeed, some people argue that word processing technology makes the physical task of writing so much easier that some people toss self-discipline to the electrons and hedonistically indulge themselves by larding their prose with everything but the kitchen sink. Conversely, the "per-

fectionists" turn into digital Flauberts, writhing in agony over which comma should go where and if that semicolon is really the best way to go.

Some products, used judiciously, aid the process of writing correctly and with good grammar, but nothing can stop the rambling author from rambling and run-ons and going on and on.

Products such as RightWriter (Cue Software), Grammatik (Reference Software), Correct Grammar (Lifetree Software), and Avalanche's Proof Positive (Avalanche Development Co.) rate documents for readability. They also provide suggested changes, to be taken with large doses of salt, of course.

These packages use readability scores to rate the document as appropriate for a particular reading grade level. A few readability indexes are widely recognized. Chief among these are the Flesch-Kincaid Score and the Fog Index.

According to the RightWriter (a grammar checker) manual: [23]

The Flesch-Kincaid formula is the United States Government Department of Defense standard (DOD MIL-M-38784B). The government requires its use by contractors producing manuals for the armed services. The Readability Index is equivalent to the Overall Reading Grade Level (OGL) for the document.

Grade Level = (.39 x ASL) + (11.8 x ASW) - 15.59.

ASL = average sentence length (# of words / # of sentences).

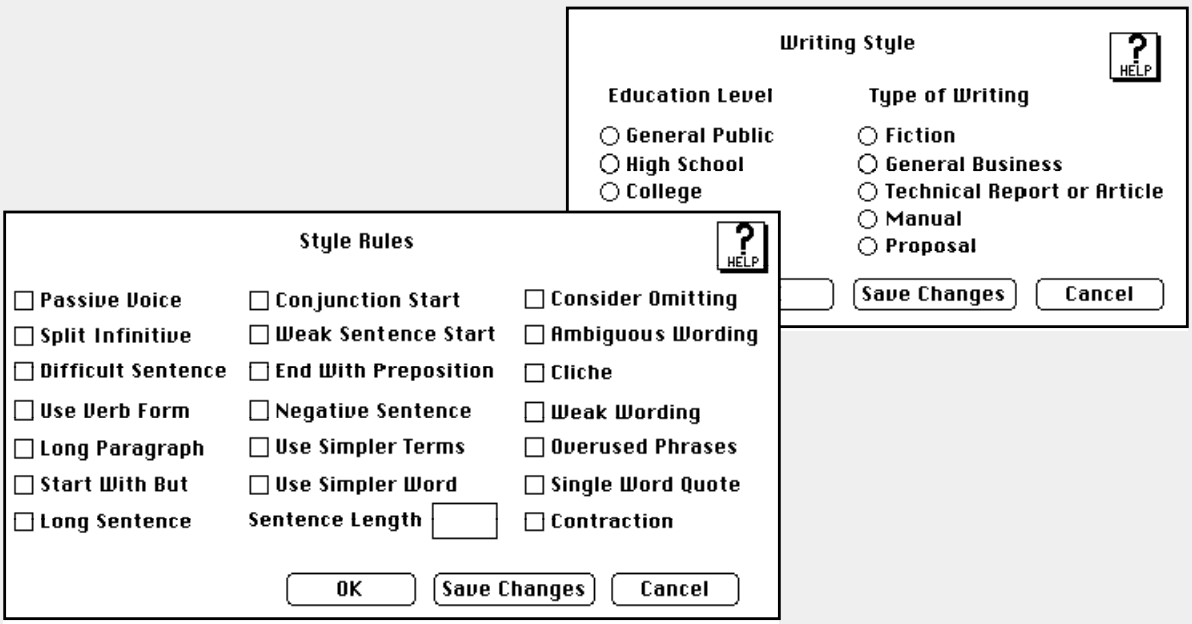
ASW = average # of syllables / word (# of syllables / # of words).

A good range is 6-10.

Grammar checker systems can generate reports about average sentence and paragraph length, the use of passive voice, the use of jargon, and other writing aspects.

RIGHTwriter® Options

The following illustrations show some of the stylistic analysis options available using RIGHTwriter® 3.1 on the Macintosh.



AT&T sells a writing tool called WWB, the Writer's Workbench software, that runs under the UNIX operating system. It's an interesting collection of utilities, that helps analyze writing style and suggests changes to fix grammatical problems. It can look for problems with punctuation, sentence length, readability, split infinitives, and overall organization. It even has a utility to compare your language style with that of another document, facilitating consistency over large numbers of documents.

1 • 3 • 2 Random Writing Tools

Aside from the various grammatical aids previously mentioned, spelling checkers are certainly the most frequently used writing tool. Spell checkers vary from ones that simply list the words not found in a dictionary to ones that make suggested corrections. The better spell checkers can work with several dic-

A Little Spelling Checker Humor

Some spelling checkers not only find spelling errors but also suggest replacements. On an electronic mailing list that discusses FrameMaker issues, there was even a series of messages about amusing word replacements. Some of the suggested replacements were bomb for IBM, salivation for Xyvision, masochist for Massachusetts and the hands–down winner—replacing Interleaf (a prime competitor) with FrameMaker.

On-line Quotes

Writing with the capability of searching for quotes can be amusing. Let's say you're preparing a presentation about writing books. A search for quotes containing the words "write" and "book" using Microsoft Bookshelf yields:

"A bad *book* is as much of a labor to *write* as a good one; it comes as sincerely from the author's soul." — Aldus Huxley

NIFTY TOOL!!

tionaries. The spell checker may be able to use a general dictionary, a site-wide (organization) dictionary, one for a user and one for the particular document.

Most of the widely used word processing packages provide or work with a built-in thesaurus. These are always useful when searching for that hard–to–think–of–word, utterance, expression, maxim, term, slogan, verbiage, declaration, idiom, phrase, remark, statement, comment, and so on.

One innovative writing tool introduced way back in 1987 is the Microsoft Bookshelf. It was one of the first serious mass market CD-ROMs and was aimed at writers. The storage capacity of the CD-ROM enabled the Bookshelf to contain eleven reference books and information data sets. Among these were *The American Heritage Dictionary, Roget's II: Electronic Thesaurus, Bartlett's Familiar Quotations, The Chicago Manual of Style* and the *U.S. ZIP Code Directory.* The combination of these reference materials in the context of a PC and a word processor is a powerful tool.

Budding poets can also start to compute. The "Rhymer" from WordPerfect Corporation is a rhyming dictionary available for use with WordPerfect on PCs, you see. One can search for words by a number of phonetic characteristics. Act like a bloodhound and search for a sound; it will simply astound, not confound. Just imagine the possibilities of rhyming for searched quotes with words found in the thesaurus! Onward writers—now you have as many tools to abuse as graphic designers do!

1 • 4 The Engineered View

Documents are complex objects. Let's now examine the document as an object composed of a variety of pieces that must be "engineered" together.

Often, the only time all pieces of a project come together is when the final report is due. All the information gathered from a variety of sources must be assembled into a coherent, deliverable product. Most likely, many people contribute to the final report. Their individual idiosyncratic uses of publishing tools must be integrated into a consistent product.

Data created by spreadsheets or images from drawing tools are also often included in completed documents. The assembly of all these components brings us to the topic of the compound document.

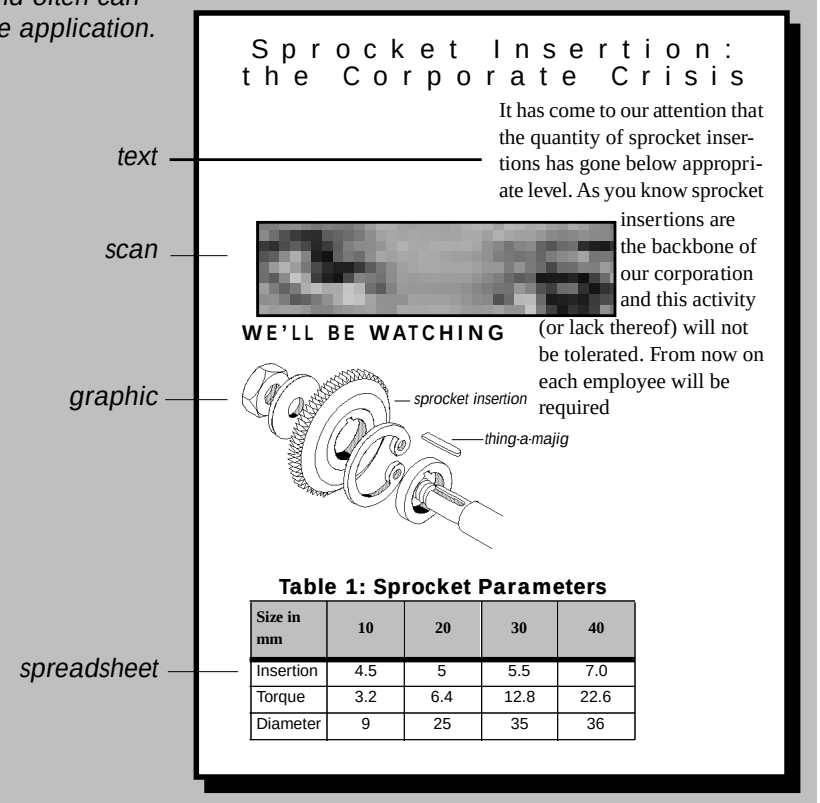
1 • 4 • 1 Compound Document

The compound document, as its name suggests, is a document composed of many parts. These parts may originate from vastly different systems and exist in many different formats. From a technical standpoint, the integration of these pieces into a coherent whole is a formidable task. Each part must be integrated seamlessly into what appears to be a single consistent document. Even more difficult is the often necessary requirement to go back to the original system that created the data. For example, a spreadsheet, in order to edit the data.

Electronically created compound documents resemble information quilts patched together from a variety of information sources. Information created for one purpose in one particular system may be used in several systems. The information may also be used for a purpose other than that for which it was intended. Documents created with such information can quickly become impossible to maintain and update.

The original data sources become an integral part of the creation process, and great care must be exercised to maintain those data sources for future versions of the document. Text, graphics, and scanned photos may be assembled for one purpose and later reassembled for another. Document content may be reused. If proper care is taken of all the various data sources, the information can be reused. Reusing the content allows an organization to profit from the publication of the content again and again.

A prototypical compound document that integrates information from many sources. The user can select each element and often can manipulate the data in its native application.

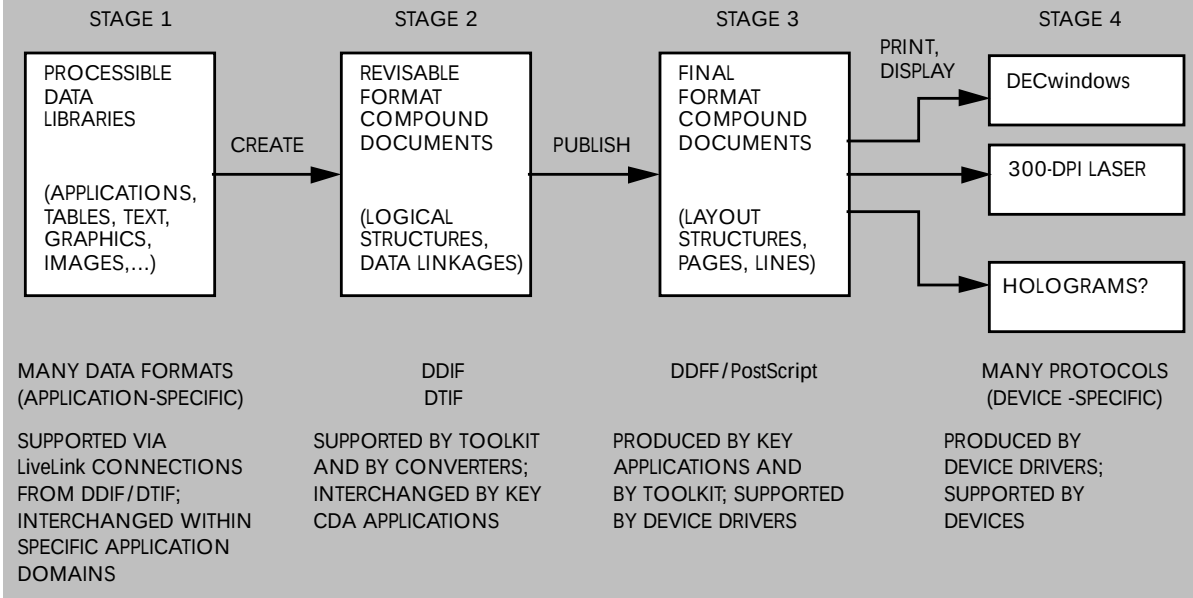


Both IBM and DEC have ongoing software projects that address the challenge of compound documents. IBM's MO:DCA (Mixed Object Document Content Architecture) is a combination compound document and object architecture. DEC's CDA (Compound Document Architecture) is a new system resembling the philosophical approach of ODA. (For more information on the Office Document Architecture standard please see *Section 3 • 4 ODA* in *Chapter 3 • Document Standards*.)

DEC's implementation approach is to provide developers with a CDA toolkit. The toolkit will give CDA developers a jump–start in the creation of applications compliant with the CDA.

CDA Document Processing Model

This illustration shows the four stages of document processing defined by the CDA. Stage 2, the revisable format, is where most editing and content are added.[24]



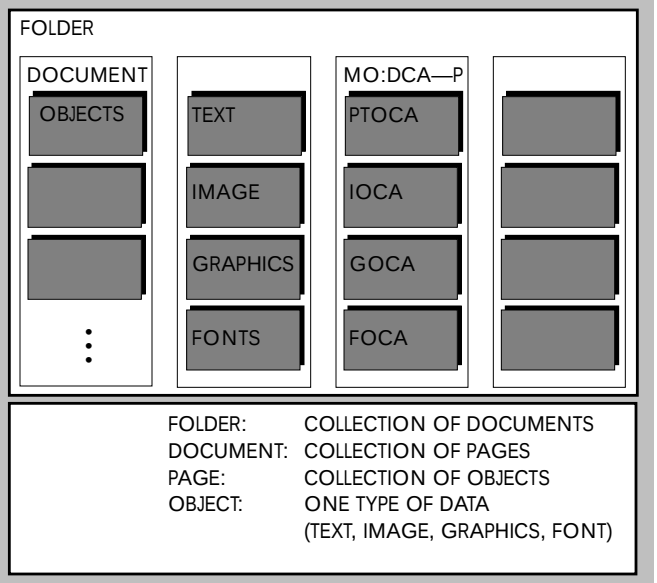
IBM's MO:DCA will have many *content architectures* for the various component parts that a document may include. IBM is working on a composite editor that will allow graphics to be placed in text; however, the graphics cannot be modified. The distinction between a composite and compound editor lies in the ability to modify component items.

The initial introduction of the MO:DCA technology will be an electronic mail product: a *correspondence processor.* Three other pieces of the MO:DCA are the IOCA (Image Object Content Architecture), GOCA (Graphics Object Content Architecture), and PTOCA (Presentation Text Object Content Architecture).[25] It is clear that the philosophical approach taken by IBM is similar to the ODA approach. Each provides an

architecture for different types of data. This bodes well for the creation of quality document exchange between systems that implement these formats. This makes perfect sense, as IBM has declared its full support of ODA and is an active sponsor of the Open Document Architecture Consortium. [26]

A document is a collection of pages that can include text, graphics, and images. A very important consideration in image work is data-stream selection. Image data are not easy to convert to alphanumeric data. Also, because of the large size of the databases that enterprises are building, data conversion may be impractical. These consideration are illustrated in the figure.

"The IBM System Application Architecture (TM) (SAA(TM)) document data stream is Mixed Object Document Content Architecture-Presentation (MO:DCA-P). MO:DCA-P is a carrier data stream that consists of objects and the layout information that specifies how a document is to be printed or displayed. Objects may be in line as part of the document, or they may reside in an external library.[27] The text data stream is Presentation Text Object Content Architecture (PTOCA), and the font data stream is Font Object Content Architecture (FOCA). Graphics, such as for annotating an image, could be represented with Graphics Object Content Architecture (GOCA)."



1 • 4 • 2 Active Documents

The various architectural approaches discussed in the previous section will permit the creation of new types of document processing. One new type is the **active document**. A number of publishing systems already tout this capability, but may call it different

things. For example, a pie chart of data from a spreadsheet, included in a document, may update itself when the spreadsheet changes. In another case, a paragraph just rewritten may initiate an electronic mail message to a manager, informing the manager of the change and requesting approval. The document is no longer a passive object; it's doing things. The notion of a document with active components is another step in the direction of a totally integrated information environment.

Several technologies are available for interprocess and interapplication communication. Publishing systems approach the problem of application communications in several ways. Ultimately, the publishing system depends on the services provided by the operating system. Most operating systems provide some mechanism for interapplication communications, and these mechanisms are exploited by some of the publishing systems. For example, on MS-DOS platforms running WINDOWS, a facility called OLE (Object Linking and Embedding) is used by MS Word for Windows to include "live" EXCELL spreadsheets. The Macintosh's System 7 operating system has a "Publish and Subscribe" facility for interapplication communication. Interleaf and FrameMaker on UNIX platforms use RPC (Remote Procedure Calls) to allow an AutoCad drawing in a document to be linked to the AutoCad application.

Interleaf's **active document technology** is one of the more ambitious implementations of the active document approach. Document sections can behave in certain ways and take various actions. For example, a document can be directed to send e-mail to various managers for approval before permission is granted for the public to view the document. This feature could prove invaluable to organizations that require complex configuration management of documents, because documents are just one portion of an engineering effort. For example, the production of an airplane must correspond accurately to the various designs and tests of the airplane. The ability to embed "intelligence" into documents is an interesting approach to the configuration management prob-

lem. For more discussion on this topic see *Section 6 • 2 Configuration Management* in *Chapter 6 • Document Management*.

1 • 5 The Database View

Let's move now to an examination of the relationship of documents to databases. Documents can relate directly to databases in two main ways. First is the report or simple printout of a database. This is known as database publishing. Second, and more interesting, is the use of a database to hold the document content. The various components that comprise a document can be placed into a database. The database can be queried by the publishing system and out pops the printed pages (ahhh... if only it was that simple). These types of systems are possible today; however, there are no hard and fast rules for accomplishing such implementations. Each organization's needs and requirements must be carefully analyzed, and no one solution will fit everyone's needs.

Reusing a document's components is becoming increasingly possible. Reuse is possible only if you can identify and reassemble pieces of content. Mechanisms to break apart the original document into meaningful component parts can be developed using standards and well-defined recommended practices.

Sometimes the seemingly simple task of finding the relevant material may in fact be the most difficult aspect of reusing content material. Document repositories must be created with appropriate key words or embedded tagging mechanisms to enable meaningful retrieval. Electronic imaging systems, which scan reams of documents and store the images on optical disks as a replacement for microfilm, are a growing industry. Without sufficient tagging, these systems are a small step forward from microfilm technology. An image of a page without the means of asking for information about what is on the page is no better then a picture. About the only savings is physical storage space (which may in fact be significant for an organization).

1 • 5 • 1 Database Publishing

An extension of report–generation capabilities, which has been around for many years, database publishing adds a level of integration between the database and the publishing system. A report is one visual representation of a database. Slick publications produced by pouring data into visual templates might be another representation of the same database.

Database publishing tools allow you to choose particular fields in a database for printing. Particular styles can be applied to these fields and selectively printed. Such tools are invaluable for catalogs with thousands or hundreds of thousands of entries, such as a yellow–pages directory and parts catalogs, which must be updated regularly.

Information from inside the database is extracted and combined with the publishing system to produce good–looking documents, not simply printouts.

The most common form of database publishing involves **merge** facilities. A merge facility combines regularly structured information, with a template document. For example, a list of names and addresses, one per line with tab separators, might be combined with a form letter that contains special codes that indicate, to the document processor, when to insert data from the data file.

The Vietnam Memorial

One of the most interesting database publishing efforts to date was the "printing" of the names on the Vietnam Memorial in Washington, D.C. The memorial contains the names of all the U.S. citizens killed in the Vietnam war. The names are etched in stone in chronological order, but within any particular day, the names are in alphabetical order. Data Development, Inc., of Palm City, Florida, was responsible for this painstaking work. This project also teaches a lesson about the quest for total accuracy. Obviously, accurate spelling of the names was of utmost importance. All 50,000 names were proofread six times. Six errors still escaped.

Merging Data

Most word processor systems provide the ability to merge data with template documents. Typically, this is called mail merge or simply a merge facility.

NAME	BIRTHDAY
Faye	12 18 55
Joseph	11 21 87
Lizzie	08 10 84
Sandy	12 08 56
Groucho	10 02 90
Harpo	11 23 88
Chico	03 22 87
Zeppo	02 25 01
Gummo	08 22 92

data

Hello!!
{NAME}

That's right, on {BIRTHDAY} you were brought into this world!!

What better way to commemorate this event than with a twenty-piece matching set of stainless steel ninja weapons. That's right, we've got the swo
chucks, the flying
matching headbar

template

Hello!!
Joseph

That's right, on 11 21 87 you were brought into this world!!

What better way to commemorate this event than with a twenty-piece matching set of stainless steel ninja weapons. That's right we've got the swords, the nun-chucks, the flying death stars, and matching headbands.

final document

Information in a database will eventually need to be explained, summarized, and otherwise communicated to someone. A tighter link between the database (information source) and the document (information sink) aids the communication process in several ways. The information can reach the docu-

ment faster and with fewer potential errors. Translations and transcriptions of database information to documents are error–prone tasks. The more direct this process can be, the better.[28]

If you take the concepts of database publishing one step farther, you arrive at the concept of a document that functions as a *front end* to a database. The information in a document that came from a database can serve as the interface to a database. Database queries via the document and automatic updates bring the database / document connection full circle. Live link and active documents provide the technological foundations for this tight linkage.

1 • 5 • 2 Customized Publishing

If a collection of information—the content of a document—is kept in the proper type of database, publishers can reuse the content and create customized texts. Hardware and software advances are both contributing to a new publishing technology, which enables documents to be custom made for particular audiences. One particularly visible example is the college textbook. McGraw-Hill and Simon & Schuster both have projects to create and distribute customized textbooks.

A partnership between McGraw-Hill and the University of Southern California has been described as follows:

> Textbook publishers are offering new computer and printing systems that allow professors to custom-design textbooks by handpicking course materials from electronic databases stocked with traditional textbooks, magazine articles and other published information.
>
> These customized books can be printed in limited quantities by the campus bookstore and distributed to students, sometimes within hours—not weeks or months—after ordering.[29]

One enterprising Washington, D.C., based company is taking another tack to custom printing. You might call it just-in-time printing. It produces an *hourly* newspaper called “The Latest News” for people who travel the Washington to New York air shuttle.[30] Information from wire services is fed into document

processing systems and formatted right away. This approach blurs the line between printed media and radio.

The ultimate in customized publishing is represented by some of the research at MIT's Media Lab. One interesting project composes information from wire services and television news. Using a computer screen with a touch screen interface, the reader can interact with this "newspaper." Fingering topics brings articles into view. Sometimes, touching a color picture brings it to life as video. This work and other projects at the Media Lab are pointing the way to personalized interactive information sources way beyond the newspaper...but I'd still like to read it on a bus.[31]

1 • 6 Specialized Views

Full–featured document processing systems often include specialized areas that have their own mini-processors. For example, mathematical equations, tables, and flow charts are all elements that can make up a document; for each, specialized document processors are available.

Specialized document processing tools support many of the semantics needed to edit these particular elements. Embedding such knowledge in the programs allows manipulations that are more natural for the particular type of document element. For example, movement of a box in a flow chart could cause all connected lines to remain attached to the box. A table editor may allow for the insertion of rows of data with a simple command. For a more through explanation of these systems, please see *Section 2 • 1 Types of Document Processors* in *Chapter 2 • Form and Function of Document Processors*.

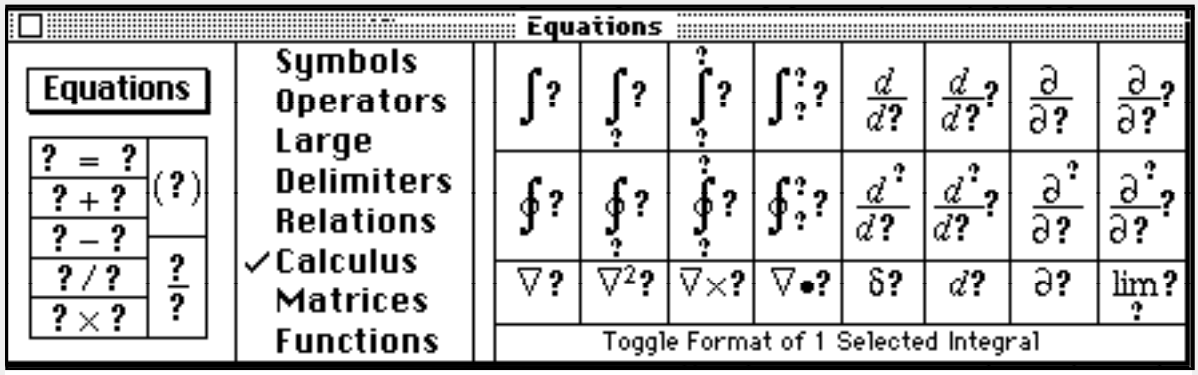
FrameMaker's Equation Editor

Within FrameMaker's publishing system is an equation editor, which actually does mathematical manipulations on the equations! While not as powerful as TeX's equation typesetting abilities, it is WYSIWYG in nature.

$$\int_0^a x^n dx = \frac{a^{n+1}}{n+1} = \frac{?}{?}\oint_?^? ?$$

an equation in the process of editing

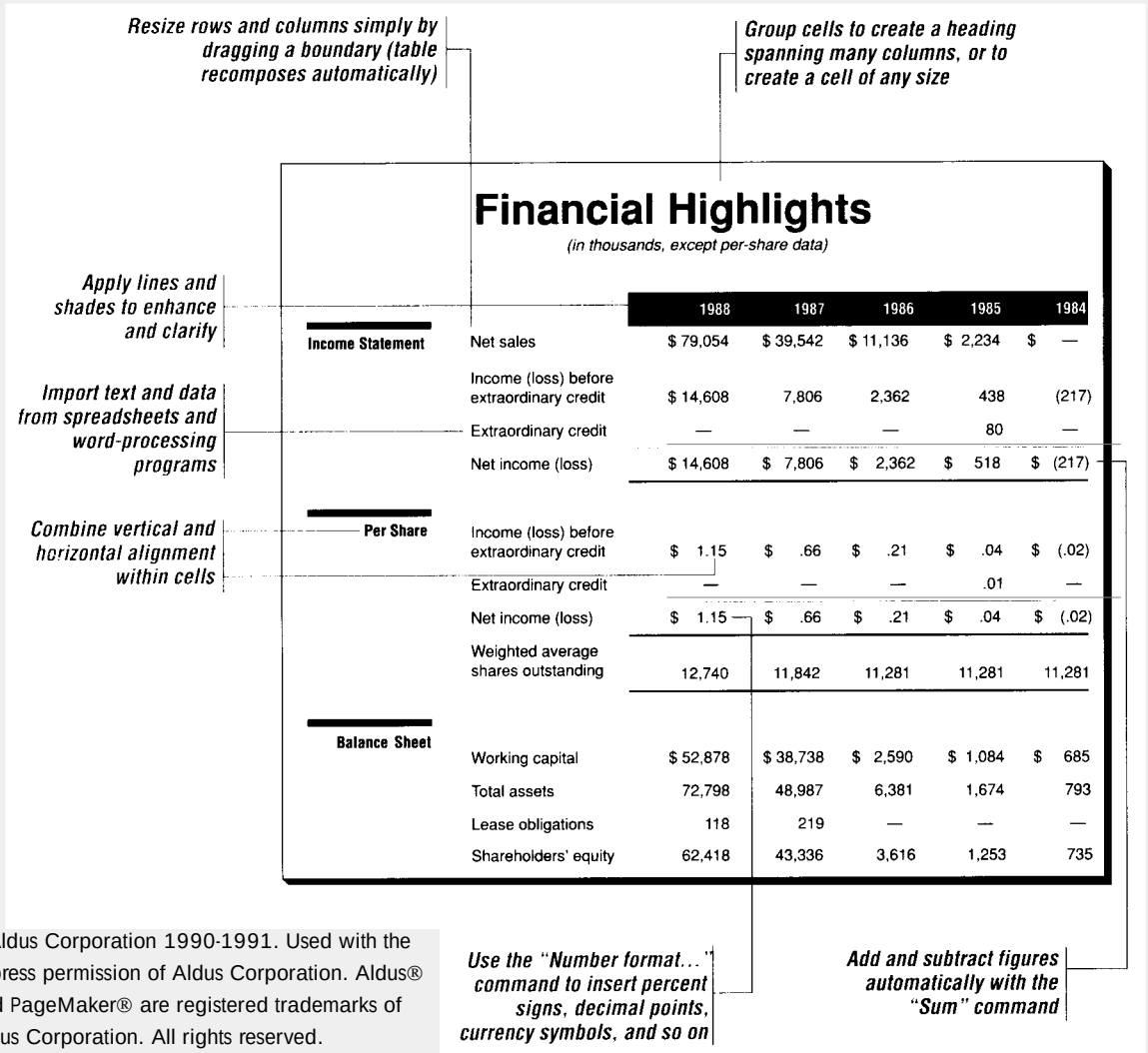
some of the controls used in the equation editor



Tables represent a particularly common document element that many systems support. Table editors exist in all kinds of electronic publishing systems, ranging from the low–end word processors such as WordPerfect, through page layout systems such as PageMaker, to the higher–end systems such as Interleaf 5.

Aldus's Table Editor

PageMaker comes complete with a separate utility for creating and editing tables. The following describes some of the types of manipulations possible with this table editor.



Financial Highlights

(in thousands, except per-share data)

		1988	1987	1986	1985	1984
Income Statement	Net sales	$ 79,054	$ 39,542	$ 11,136	$ 2,234	$ —
	Income (loss) before extraordinary credit	$ 14,608	7,806	2,362	438	(217)
	Extraordinary credit	—	—	—	80	—
	Net income (loss)	$ 14,608	$ 7,806	$ 2,362	$ 518	$ (217)
Per Share	Income (loss) before extraordinary credit	$ 1.15	$ .66	$ .21	$ .04	$ (.02)
	Extraordinary credit	—	—	—	.01	—
	Net income (loss)	$ 1.15	$ .66	$ .21	$ .04	$ (.02)
	Weighted average shares outstanding	12,740	11,842	11,281	11,281	11,281
Balance Sheet	Working capital	$ 52,878	$ 38,738	$ 2,590	$ 1,084	$ 685
	Total assets	72,798	48,987	6,381	1,674	793
	Lease obligations	118	219	—	—	—
	Shareholders' equity	62,418	43,336	3,616	1,253	735

Another interesting and growing specialized document processing field is legal document assembly. The high-end document assembly packages such as CAPS by CAPSoft or WorkForm by Analytic Legal

Programs allow Joe Lawyer to produce document templates that can be used by others on the staff. The templates are produced by answering a series of questions to direct the software to assemble the document by pulling in the correct text from its textual database. Accuracy is of the utmost importance in legal documents. One misplaced word can be the source of litigation or of numerous other complexities.[32]

Maintenance manuals used by the military raise another legal issue. These manuals usually contain lots of WARNING boxes indicating some important message. For example, applying more that 10 fps torque will cause death and destruction. It is legally mandated that the WARNING be placed before the text of the section covering that topic.

The placement of mandatory items has some interesting ramifications for on-line reading, exemplified by hypertext browsers.[34] An on-line document browsing system must be designed to display the WARNING before allowing the display of the associated text. Random browsing through the document must factor in this requirement. The trick, of course, is to do this without interfering with the flexibility of browsing and searching, which is desirable in the online document viewers.

Details Count, Oh Those Quote Marks

Even the seemingly tiniest detail can have major implications. For example, a recent *Washington Post* article titled "U.S. Appeals Court Finds Error Curbs Insurance Sales by Banks"[33] reported:

> A federal appeals court yesterday threw bankers, insurance agents and their lawyers into a tizzy by ruling that back in 1918, Congress misplaced a pair of quotation marks and accidently repealed the law that allows national banks in small towns to sell insurance....
>
> OCC (Office of the Comptroller of the Currency) attorneys argued that Congress didn't mean to take away the power, but that claim "runs up against the stubborn fact that the troublesome quotation marks are located where they are, not where the parties argue that the 64th Congress intended them to be," said Judge James L. Buckley.

Two techniques to manage complexity and avoid errors are configuration management and clear project organization.

1 • 7 Summary

There is no single correct way to look at document processing issues. Each project has unique constraints and circumstances. However, it is important to appreciate that different points of view exist and are useful.

For one project, design may be paramount, and for another, the logical structure may be critical. In the end any evaluation of a publishing system depends on what you need for a particular project.

In any evaluation of a system, half the battle is to ask good questions. The various points of view discussed in this chapter provide a useful frame of reference that will help you to ask good questions.